Chapter 7: Parameter Estimation in Time Series Models

- In Chapter 6, we learned about how to specify our time series model (decide which specific model to use).
- The general model we have considered is the $ARIMA(p, d, q)$ model.
- The simpler models like AR, MA, and ARMA are special cases of this general $ARIMA(p, d, q)$ model.
- Now assume we have chosen appropriate values of p, d, and q (possibly based on evidence from the ACF, PACF, and/or EACF plots).
- Assume that our observed time series data $Y_1, \ldots, Y_n$ follow a stationary $ARMA(p, q)$ model.
- In the case of nonstationary original data, we can assume that taking d differences has produced differenced data that displays stationarity.
- We now must estimate the unknown parameters in that stationary $ARMA(p, q)$ model.

Method of Moments Estimation

- One of the easiest methods of parameter estimation is the *method of moments* (MOM).
- The basic idea is to find expressions for the sample moments and for the population moments and equate them:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^r = E(X^r)$$

- The $E(X^r)$ expression will be a function of one or more unknown parameters.
- If there are, say, 2 unknown parameters, we would set up MOM equations for $r = 1, 2$, and solve these 2 equations simultaneously for the two unknown parameters.
- In the simplest case, if there is only 1 unknown parameter to estimate, then we equate the sample mean to the true mean of the process and solve for the unknown parameter.

MOM with AR models

- First, we consider autoregressive models.
- In the simplest case, the $AR(1)$ model, given by $Y_t = \phi Y_{t-1} + e_t$, the true lag-1 autocorrelation $\rho_1 = \phi$.
- For this type of model, a method-of-moments estimator would simply equate the true lag-1 autocorrelation to the sample lag-1 autocorrelation r_1.
- So our MOM estimator of the unknown parameter ϕ would be $\hat{\phi} = r_1$.

MOM with an $AR(2)$ model

- In the $AR(2)$ model, we have unknown parameters ϕ_1 and ϕ_2.
- From the Yule-Walker equations,

$$\rho_1 = \phi_1 + \rho_1 \phi_2 \text{ and } \rho_2 = \rho_1 \phi_1 + \phi_2$$

- In the method of moments, we will replace the true lag-1 and lag-2 autocorrelations, ρ_1 and ρ_2, by the sample autocorrelations r_1 and r_2, respectively.

MOM with an $AR(2)$ model, continued

- That gives the equations

$$r_1 = \phi_1 + r_1\phi_2 \text{ and } r_2 = r_1\phi_1 + \phi_2$$

  which are then solved for ϕ_1 and ϕ_2 to obtain

$$\hat{\phi}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2} \text{ and } \hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$$

- The general $AR(p)$ model is estimated in a similar way, with the Yule-Walker equations being used to obtain the *Yule-Walker estimates* $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$.

MOM with MA Models

- We run into problems when trying to using the method of moments to estimate the parameters of moving average models.
- Consider the simple $MA(1)$ model, $Y_t = e_t - \theta e_{t-1}$.
- The true lag-1 autocorrelation in this model is $\rho_1 = -\theta/(1+\theta^2)$.
- If we equate ρ_1 to r_1, we get a quadratic equation in θ.
- If $|r_1| < 0.5$, then only one of the two real solutions satisfies the invertibility condition $|\theta| < 1$.
- That solution is $\hat{\theta} = \left(-1 + \sqrt{1 - 4r_1^2}\right)/(2r_1)$.
- But if $|r_1| = 0.5$, no invertible solution exists, and if $|r_1| > 0.5$, then no real solution at all exists, and the method of moments fails to give any estimator of θ.

More MOM Problems with MA Models

- With higher-order $MA(q)$ models, the set of equations for estimating $\theta_1, \ldots, \theta_q$ is highly nonlinear and could only be solved numerically.
- There would be many solutions, only one of which is invertible.
- In any case, for $MA(q)$ models, the method of moments usually produces poor estimates, so it is not recommended to use MOM to estimate MA models.

MOM Estimation of Mixed ARMA Models

- Consider only the simplest mixed model, the $ARMA(1,1)$ model.
- Since $\rho_2/\rho_1 = \phi$, a MOM estimator of ϕ is $\hat{\phi} = r_2/r_1$.
- Then the equation

$$r_1 = \frac{(1-\theta\hat{\phi})(\hat{\phi}-\theta)}{1-2\theta\hat{\phi}+\theta^2}$$

  can be used to solve for an estimate of θ.
- This is a quadratic equation in θ, and so we again keep only the invertible solution (if any exist) as our $\hat{\theta}$.

MOM Estimation of the Noise Variance

- We still must estimate the variance σ_e^2 of our error component.
- For any model, we first estimate the variance of the time series process itself, $\gamma_0 = var(Y_t)$, by the sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_t - \bar{Y})^2$$

- Then we can take advantage of known relationships among the parameters in our specified model to obtain a formula for $\hat{\sigma}_e^2$.

Formulas for MOM Noise Variance Estimators in Common Models

- For $AR(p)$ models, $\hat{\sigma}_e^2 = (1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2 - \cdots - \hat{\phi}_p r_p)s^2$.
- For the $AR(1)$ model, this reduces to $\hat{\sigma}_e^2 = (1 - r_1^2)s^2$.
- For $MA(q)$ models,

$$\hat{\sigma}_e^2 = \frac{s^2}{1 + \hat{\theta}_1^2 + \hat{\theta}_2^2 + \cdots + \hat{\theta}_q^2}.$$

- For $ARMA(1,1)$ models,

$$\hat{\sigma}_e^2 = \frac{1 - \hat{\phi}^2}{1 - 2\hat{\phi}\hat{\theta} + \hat{\theta}^2} s^2.$$

MOM Estimation in Some Simulated Time Series

- The course web page has R code to estimate the parameters in several simulated AR, MA, and ARMA models.
- The estimates of the AR parameters are good, but the estimates of the MA parameters are poor.
- In general, MOM estimators for models with MA terms are inefficient.

MOM Estimation in Some Real Time Series (Hare data)

- On the course web page, we see some estimation of parameters for real time series data.
- For the Canadian hare data, we employ a square-root transformation and select an $AR(2)$ model:

$$(\sqrt{Y_t} - \mu) = \phi_1(\sqrt{Y_{t-1}} - \mu) + \phi_2(\sqrt{Y_{t-2}} - \mu) + e_t$$

- Note that because the mean of the process is not zero, we initially subtract off $\mu = E(\sqrt{Y_t})$ throughout.
- Using the method of moments, we estimate the unknown parameters μ, ϕ_1, and ϕ_2 (see R example).
- The final estimated model is

$$(\sqrt{Y_t} - 5.82) = 1.1178(\sqrt{Y_{t-1}} - 5.82) - 0.519(\sqrt{Y_{t-2}} - 5.82) + e_t$$

with estimated noise variance 1.97.

MOM Estimation in Real Time Series (Oil price data)

- For the Oil price data, we select an $MA(1)$ model for the differences of the logged oil prices:

$$(\nabla \log Y_t - \mu) = e_t - \theta e_{t-1}$$

- We again subtract off $\mu = E(\nabla \log Y_t)$ throughout to account for the fact that the real data may not have mean zero.
- Using the method of moments, we estimate the unknown parameters μ and θ (see R example).
- The final estimated model is

$$(\nabla \log Y_t - 0.004) = e_t + 0.222 e_{t-1}$$

with estimated noise variance 0.00686.

- Based on the standard error of the estimate of μ (see formula on page 28), it could be argued that the value of 0.004 is not significantly different from 0, so we could drop this 0.004 from the final model.

Least Squares Estimation

- Since method-of-moments performs poorly for some models, we examine another method of parameter estimation: *Least Squares*.
- We first consider autoregressive models.
- We assume our time series is stationary (or that the time series has been transformed so that the transformed data can be modeled as stationary).
- To account for the possibility that the mean is nonzero, we subtract μ from each observation and treat μ as a parameter to be estimated.

LS Estimation for the $AR(1)$ Model

- Consider the mean-centered $AR(1)$ model:

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$$

- The least squares method seeks the parameter values that minimize the sum of squared differences:

$$S_c(\phi, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2$$

- This criterion is called the *conditional sum-of-squares function* (CSS).

LS Estimation of μ for the $AR(1)$ Model

- Taking the derivative of CSS with respect to μ, setting equal to 0 and solving for μ, we obtain the LS estimator of μ:

$$\hat{\mu} = \frac{1}{(n-1)(1-\phi)}\left[\sum_{t=2}^{n} Y_t - \phi \sum_{t=2}^{n} Y_{t-1}\right]$$

- For large n, this $\hat{\mu} \approx \bar{Y}$, regardless of the value of ϕ.

LS Estimation of ϕ for the $AR(1)$ Model

- Taking the derivative of CSS with respect to ϕ, setting equal to 0 and solving for ϕ, we obtain the LS estimator of ϕ:

$$\hat{\phi} = \frac{\sum_{t=2}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=2}^{n}(Y_{t-1} - \bar{Y})^2}$$

- This estimator is almost identical to r_1: it's just missing one term in the denominator, $(Y_n - \bar{Y})^2$.
- So, especially for large n, the LS and MOM estimators are nearly identical in the $AR(1)$ model.
- In the general $AR(p)$ model, the LS estimators of μ and of $\phi_1, \ldots, \phi_p$ are approximately equal to the MOM estimators, especially for large samples.

LS Estimation for Moving Average Models

- Consider now the $MA(1)$ model:

$$Y_t = e_t - \theta e_{t-1}$$

- Recall that this can be written as

$$Y_t = -\theta Y_{t-1} - \theta^2 Y_{t-2} - \theta^3 Y_{t-3} - \cdots + e_t.$$

- So a least squares estimator of θ can be obtained by finding the value of θ that minimizes

$$S_c(\theta) = \sum [Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \theta^3 Y_{t-3} + \cdots]^2$$

- But this is nonlinear in θ, and the infinite series causes technical problems.

LS Estimation for Moving Average Models

- Instead, we proceed by conditioning on one previous value of e_t. Note that
$$e_t = Y_t + \theta e_{t-1}$$
- If we set $e_0 = 0$, then we have the set of recursive equations $e_1 = Y_1$, $e_2 = Y_2 + \theta e_1, \ldots, e_n = Y_n + \theta e_{n-1}$.
- Since we know $Y_1, Y_2, \ldots, Y_n$ (these are the observed data values) and can calculate the $e_1, e_2, \ldots, e_n$ recursively, the only unknown quantity here is θ.
- We can do a numerical search for the value of θ (within the invertible range between -1 and 1) that minimizes $\sum (e_t)^2$, conditional on $e_0 = 0$.
- A similar approach works for higher-order $MA(q)$ models, except that we assume $e_0 = e_{-1} = \cdots = e_{-q} = 0$ and the numerical search is multidimensional, since we are estimating $\theta_1, \ldots, \theta_q$.

LS Estimation for ARMA Models

- With the *ARMA*(1, 1) model:

  $$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1},$$

  we note that

  $$e_t = Y_t - \phi Y_{t-1} + \theta e_{t-1}$$

  and minimize $S_c(\phi, \theta) = \sum_{t=2}^{n} e_t^2$; note that the sum starts at $t = 2$ to avoid having to choose an "initial" value Y_0.
- With the general *ARMA*(*p*, *q*) model, the procedure is similar, except that we assume $e_p = e_{p-1} = \cdots = e_{p+1-q} = 0$, and we estimate $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$.
- For *large samples*, when the parameter sets yield invertible models, the initial values for $e_p, e_{p-1}, \ldots, e_{p+1-q}$ have little effect on the final parameter estimates.

Maximum Likelihood Estimation

- On the other hand, for small to moderate sample sizes (and for stochastic seasonal models), assuming $e_p = e_{p-1} = \cdots = e_{p+1-q} = 0$ can greatly affect the final parameter estimates, which is undesirable.
- In those cases, rather than using least squares, it may be advantageous to use *maximum likelihood* (ML) estimation.
- An advantage of ML estimation is that it uses all of the information in the data (not just the first few moments as in MOM).
- Also, many large-sample results are known about the sampling distribution of ML estimators.
- A disadvantage of ML estimation is that we must assume the form of the joint probability distribution of the time series process.

Maximum Likelihood in Time Series Models

- The *likelihood function* is the joint density function of the data, but treated as a function of the unknown parameters, given the observed data $Y_1, \ldots, Y_n$.
- For the models we have studied, the likelihood L is a function of the ϕ's, θ's, μ, and σ_e^2, given the observed $Y_1, \ldots, Y_n$.
- The maximum likelihood estimates (MLEs) are the values of the parameters that maximize this likelihood function, i.e., that are the "most likely" parameter values given the data we actually observed.

Maximum Likelihood in the $AR(1)$ Model

- In the $AR(1)$ model with an unknown but constant mean, the parameters we must estimate are ϕ, μ, and σ_e^2.
- To perform ML estimation in the $AR(1)$ model, we must assume a distribution for our data.
- The typical assumption is that the $\{e_t\}$ in the $AR(1)$ model are iid $N(0, \sigma_e^2)$ random variables.
- Under this assumption, the likelihood function (details are given on page 159) is:

$$L(\phi, \mu, \sigma_e^2) = (2\pi\sigma_e^2)^{-n/2}(1-\phi^2)^{1/2} \exp\left[-\frac{1}{2\sigma_e^2} S(\phi, \mu)\right]$$

where
$S(\phi, \mu) = \sum_{t=2}^{n}[(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 + (1-\phi^2)(Y_1 - \mu)^2$.

MLE's in the $AR(1)$ Model

- This $S(\phi, \mu)$ is called the *unconditional sum-of-squares* function.
- We must find estimates $\hat{\phi}$, $\hat{\mu}$, and $\hat{\sigma}_e^2$ that maximize the likelihood function (in practice, we typically maximize the log-likelihood function, which produces equivalent estimates).
- The estimator of the noise variance σ_e^2, in terms of the other estimates, is

$$\hat{\sigma}_e^2 = \frac{S(\hat{\phi}, \hat{\mu})}{n}.$$

- Note that dividing by $n - 2$ rather than n produces a less biased estimator, but for large sample sizes, this makes little practical difference.

MLE's in the $AR(1)$ Model

- We still need to estimate ϕ and μ.
- Comparing the *unconditional sum-of-squares* function to the *conditional sum-of-squares* function we saw earlier, note that $S(\phi, \mu) = S_c(\phi, \mu) + (1 - \phi^2)(Y_1 - \mu)^2$, so for large sample sizes, $S(\phi, \mu) \approx S_c(\phi, \mu)$.
- This implies that our ML estimates of ϕ and μ will be very similar to the LS estimates, at least for large sample sizes.
- The likelihood function for general ARMA models is more complicated, but ML estimates can usually be found in these models.
- In practice, for AR models, MA models, or general ARMA or ARIMA models, we can often find either the LS estimates or the ML estimates easily using R.

Properties of the Estimators

- Recall that LS estimators and ML estimators become approximately equal for large samples.
- So the large-sample properties of LS estimators and ML estimators are identical for basic ARMA-type models.
- For large n, these estimators are approximately unbiased and normally distributed.
- Note: For AR models, MOM estimators have identical *large-sample* properties as LS and ML estimators.
- But for models with MA terms, MOM estimators have *poor performance* and should not be used!
- For some common models, variance and correlation results for the estimators are given on page 161.

Properties of the Estimators in $AR(1)$ and $MA(1)$ models

- For example, for the $AR(1)$ model, $var(\hat{\phi}) \approx (1 - \phi^2)/n$, and for the $MA(1)$ model, $var(\hat{\theta}) \approx (1 - \theta^2)/n$.
- Clearly, the variance of the estimator decreases (i.e., the precision improves) as n increases.
- For the $AR(1)$ model, the variance of the estimator $\hat{\phi}$ will be low when the true ϕ is near 1.
- For the $MA(1)$ model, the variance of the estimator $\hat{\theta}$ will be low when the true θ is near 1.

Parameter Estimation with Some Simulated Time Series

- See the course web page for R examples for parameter estimation for two different simulated AR(1) series, each with $n = 60$, using the MOM, LS, and ML methods.
- See the course web page for R examples for parameter estimation for a simulated AR(2) series, with $n = 120$, using the MOM, LS, and ML methods.
- See the course web page for R examples for parameter estimation for a simulated ARMA(1,1) series, with $n = 100$, using the LS and ML methods (why not MOM here?).
- For these sample sizes, the various methods perform similarly in terms of their accuracy of estimation.
- With smaller sample sizes, the methods may produce more different results.

Parameter Estimation with the Color Property Time Series

- For the color property time series, we had specified an $AR(1)$ model.
- The R examples show the estimation of ϕ using the MOM, LS, and ML methods (note $n = 35$ here).
- From the ML estimate, the estimated $AR(1)$ model would be

$$Y_t = 0.57Y_{t-1} + e_t$$

- Since the mean of the color property series is clearly not zero, it is better to estimate a *mean-centered* version of the model, and using the `arima` function tells us that $\hat{\mu} = 74.33$:

$$(Y_t - 74.33) = 0.57(Y_{t-1} - 74.33) + e_t$$

  where the noise variance is estimated to be 24.83.
- Since $\rho_k = \phi^k$ for an $AR(1)$ process, we see that the autocorrelations will be positive for any lag, but will die off as the lag k increases.

Parameter Estimation with the Hare Abundance Time Series

- For the Canadian hare abundance data, recall that we will take the square root of the original abundance values.
- In the previous MOM example, we modeled the data with an $AR(2)$ model, but here we choose an $AR(3)$ model, which may be more appropriate based on the PACF.
- The R examples show the estimation of ϕ_1, ϕ_2, ϕ_3 and μ (as well as σ_e^2) using the MOM, LS, and ML methods (note $n = 31$ here).
- The final estimated model (from the ML estimates) is:

$$(\sqrt{Y_t} - 5.69) = 1.052(\sqrt{Y_{t-1}} - 5.69) - 0.229(\sqrt{Y_{t-2}} - 5.69) - 0.393(\sqrt{Y_{t-3}} - 5.69) + e_t$$

with estimated noise variance 1.066.

Parameter Estimation with the Hare Abundance Time Series (Continued)

- From the standard errors of the estimates, the lag-2 coefficient does not appear significantly different from zero.
- So we could optionally drop the lag-2 term and refit the AR model with only the lag-1 and lag-3 terms.

Parameter Estimation with the Oil Price Time Series

- Our earlier analysis specified an $MA(1)$ model for the differences of the logged oil prices.
- The R example shows the estimation of θ using several methods.
- Again, the method of moments is not recommended for the $MA(1)$ model.

Parameter Estimation with Other Time Series

- See the R examples on parameter estimation for several other data sets:
- We estimate the parameters of an $AR(2)$ model for the recruitment data.
- We estimate the parameters of an $MA(1)$ model for the differenced logged varve data.
- Either an $AR(1)$ model or an $MA(2)$ model seems to fit the differences of the logged GNP data well.

Large-sample Inference about the Model Parameters

- When the model parameters are estimated by the ML method, then the ML estimators are approximately normally distributed when n is large.
- So we can use normal-based inference to get, say, confidence intervals for the true values of the parameters.
- For example, it may be of interest to know whether 0 is a plausible value of some parameter.
- For large samples, a $(1-\alpha)100\%$ CI for a parameter takes the form:

$$\text{estimate } \pm (z_{\alpha/2})(\text{estimated standard error})$$

- For example, in an $AR(1)$ model, a 95% CI for ϕ is:

$$\hat{\phi} \pm 1.96\sqrt{(1-\hat{\phi}^2)/n}$$

- For example, in an $MA(1)$ model, a 90% CI for θ is:

$$\hat{\theta} \pm 1.645\sqrt{(1-\hat{\theta}^2)/n}$$

Small-sample Inference about the Model Parameters

- The ML estimators are not necessarily approximately normally distributed when n is small.
- So when n is small, we can use a more general approach, bootstrap-based inference, to get confidence intervals for the true values of the parameters.
- Section 7.6 gives details about bootstrap intervals.
- Some R examples give code for calculating 95% bootstrap CIs for ARIMA-type model parameters using four different methods; note that Method IV makes the fewest assumptions about the error distribution.
- The bootstrap method also makes it possible to construct CIs about relevant functions of the model parameters.